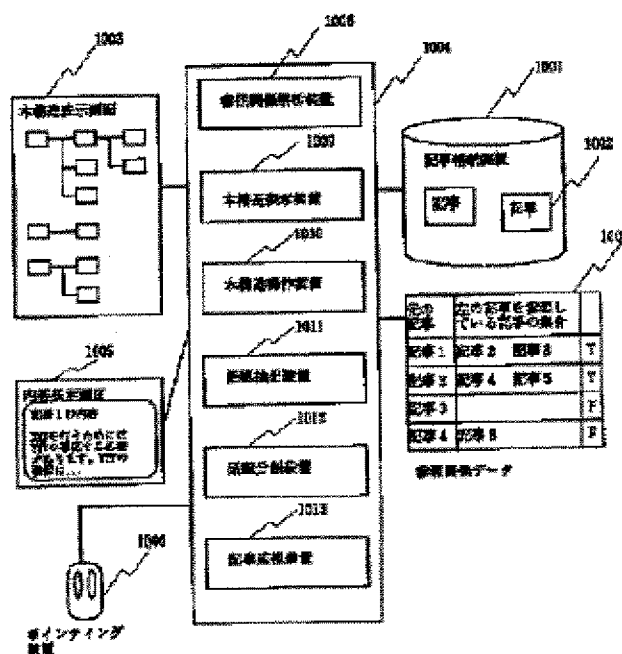


Publication number: JP9106331
Publication date: 1997-04-22
Inventor: KOGA AKIHIKO; TSUJI HIROSHI
Applicant: HITACHI LTD
Classification:
- international: *G06F13/00; G06F3/048; G06F3/14; G06F17/30; H04L12/56; G06F13/00; G06F3/048; G06F3/14; G06F17/30; H04L12/56; (IPC-1-7): G06F3/14; G06F13/00; G06F17/30; H04L12/56*
- European:
Application number: JP19950263841 19951012
Priority number(s): .JP19950263841 19951012

Report a data error here

PROBLEM TO BE SOLVED: To classify articles for every unity of stories by storing the sets of articles, analyzing the reference relation, calculating the feature functions of the articles on the tree structure display screen of a link structure and cutting a link between nodes having a large change. **SOLUTION:** An article storage device 1001 stores articles, a reference relation analysis device 1008 analyzes the reference relation between plural articles 1002 and a tree structure display device 1009 displays this reference relation data 1007 on a tree structure display screen 1003. At this stage, the articles are displayed by a node and a reference relation is displayed by the link from an article to be referred to a referred article. A tree structure operation unit 1010 makes the descendant node of the node designated by a pointing device 1006 unseen from this tree structure display screen 1003. Further, when the difference of feature vector showing the contents of articles with each other between nodes which are in a parental relation is a prescribe value or more, a topic dividing device 1012 cuts the link between these nodes and makes a separate tree as an independent topic.



6/7/2007

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-106331

(43)公開日 平成9年(1997)4月22日

(51)Int.Cl. ⁶	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 F 3/14	3 4 0		G 0 6 F 3/14	3 4 0 C
	13/00	3 5 1		13/00
	17/30			15/413
H 0 4 L 12/56		9466-5K	H 0 4 L 11/20	1 0 2 B

審査請求 未請求 請求項の数9 O L (全 15 頁)

(21)出願番号 特願平7-263841

(22)出願日 平成7年(1995)10月12日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 古賀 明彦

神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(72)発明者 ▲辻▼ 洋

神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(74)代理人 弁理士 小川 勝男

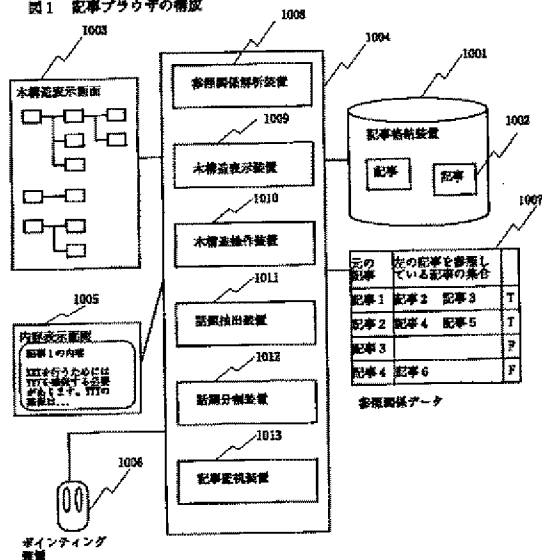
(54)【発明の名称】 電子掲示板ブラウザ

(57)【要約】

【課題】 電子掲示板には、複数の人間から有用な情報を含む記事が多数投稿されるが記事の個数が多く、どんな話題が議論されているか把握しにくい。本発明では、記事の集まりを、記事間の参照関係を使って話題毎に分割して表示、検索する手段を与える。

【解決手段】 記事の集合を蓄積する装置と記事間の参照関係を解析する装置とリンク構造を木で表示する木構造表示画面と、画面で指定された記事の内容を表示する内容表示画面、大きな木の記事の特徴関数を計算し、変化の大きなノード間でリンクを切断する話題分割装置からなる。

図1 記事ブラウザの構成



【特許請求の範囲】

【請求項1】記事の集合を納める記事蓄積装置を持ち、該記事はヘッダ部分と内容部分からなり該記事のヘッダ部分には、その記事を識別するための識別名と、その記事が別の記事を参照して書かれた場合には、参照している記事の識別名を含み、内容部分にはテキストデータを含み、該記事の内容部分を表示する内容表示画面からなる電子掲示板ブラウザにおいて、該記事の間の参照関係を表示する木構造表示画面と該木構造表示画面の上に表示されている該記事をユーザが指定するためのポインティング装置と、該記事蓄積装置内部の記事を調べて参照関係を表す参照関係データを作成する参照関係解析装置、該参照関係解析装置が作成したデータを木構造表示画面に該記事をノードで、参照している関係を参照される記事から参照した記事へのリンクで表示する木構造表示装置と、ユーザによって指定された該ノードの子孫のノードを該木構造表示画面から隠す機能を有する木構造操作装置を持ち、ユーザによって該ポインティング装置で指定された該木構造表示画面のノードに対応する記事の内容を該内容表示画面に表示することを特徴とする電子掲示板ブラウザ。

【請求項2】請求項1の電子掲示板ブラウザにおいて、複数の子供を持つノードにおいて、子供が終端ノードであるとき、他の子供で終端でないものがあれば、終端の子供を木から除去する機能からなる話題抽出装置を持つことを特徴とする電子掲示板ブラウザ。

【請求項3】請求項1の電子掲示板ブラウザにおいて、各記事のヘッダ部分に参照を表すデータがないとき、参照関係解析装置が各記事からキーワードの集合を抽出する機能を持ち、キーワードの集合の包含関係を使って記事の間のリンクを張ることを特徴とする電子掲示板ブラウザ。

【請求項4】請求項1の電子掲示板ブラウザにおいて、各記事のヘッダ部分に参照を表すデータがないとき、参照関係解析装置が各記事から他記事を参照する部分に現れやすい予め決められている文字列を検索し、その文字列の近くの文章からキーワードを抽出し、そのキーワードを含む記事を探してリンクを張ることを特徴とする電子掲示板ブラウザ。

【請求項5】請求項1の電子掲示板ブラウザにおいて、各記事に対して記事の内容を表す特徴ベクトルの蓄積し、親子関係のあるノード間で該特徴ベクトルどうしの差があらかじめ決めたある値以上であるとき、該ノード間のリンクを切る話題分割装置を持つことを特徴とする電子掲示板ブラウザ。

【請求項6】請求項5の電子掲示板ブラウザにおいて、該話題分割装置は、各記事の特徴量の差でなく、その記事のリンクに関する近傍の記事の特徴量の加重平均の差をリンクの切り離しの条件に用いることを特徴とする電子掲示板ブラウザ。

【請求項7】請求項5の電子掲示板ブラウザにおいて、該話題分割装置はリンクを切る前に切る部分を該木構造表示画面で示し、ユーザに切るかどうか聞いて確かめることを特徴とする電子掲示板ブラウザ。

【請求項8】請求項1の電子掲示板ブラウザにおいて、木の一部分を別の記事の子供に付け替える機能を持つ木構造操作装置を持つことを特徴とする電子掲示板ブラウザ。

【請求項9】請求項1の電子掲示板ブラウザにおいて、ユーザがあるノードの集合を指定して、そのノードの子孫となる記事が該記事記憶装置に入力されたら、ユーザにその記事を知らせる記事監視装置を持つことを特徴とする電子掲示板ブラウザ。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ネットワークを使って不特定多数のユーザによって電子掲示板に投稿された文書の集合を調べ、ユーザが必要な情報を検索する電子掲示板ブラウザにかかわるものである。

【0002】

【従来の技術】電子掲示板ブラウザとしては、(株)ソフトバンク発行の雑誌UNIX USER Vol2. No. 2 FEB. 1 1993 (ユニックスユーザ)、pp 115-124の記事「Whatis UNIX」とUNIX USER Vol 2. No. 3 MAR. 1 1993 pp115-125 (ユニックスユーザ)の記事「Whatis UNIX」に記載されて紹介されているGNUプロジェクト(ヌープロジェクト)が作成したGNUS(ヌース)がある。ある記事Aが別の記事Bを参照して意見を行っているときAはBをフォローアップしているという。GNUSでは、記事のフォローアップ関係を次のように表示カラムをインデントして表示し、記事と記事の関係を分かりやすくしている。

【0003】914: [aさん] 話題A

915: [bさん]

916: [aさん]

917: [cさん]

918: [dさん] 話題B

ここで、記事915は、記事914をフォローアップしており、記事915は記事916からフォローアップされている。記事917は、記事914をフォローアップしているが、記事915や記事916とはフォローアップの関係を持たない。この表示により、記事914、915、916が一つの続いた話であり、記事914、917も続いた話であることが分かる。

【0004】また、GNUSでは、記事を読んでいる最中に、その記事がフォローアップしている元の記事を検索して表示する機能や、読んでいる記事がフォローアップしている記事と同じ記事にフォローアップしている記事の集合を順に参照する機能があり、一つの記事に対するいろいろな角度からの意見を表明している記事を参照することができる。

【0005】

【発明が解決しようとする課題】GNUSでは、
 (ア) 電子掲示板にある記事の集合を参照の関係で段下げて、木として表示するので1つの話題が1つの木として表され、どのような話題が展開されるかを知る手掛かりにはなったが、1つの木が大きくなりすぎることがしばしばあり、通常の小さな画面では木の中の非常に狭い部分しか見え、電子掲示板全体でどのような話題がいくつ展開されているかを知ることは難しかった。従って、人間が把握しやすい大きさの塊で表示する手段が必要であった。

【0006】(イ) 参照で作られた木の中に含まれる記事には、全体的话题を理解するために必要なものとそうでないものがあつた。話題の理解に必要なものだけを読みみたいときに、話題と関係の薄い記事を取り去る手段がなかった。

【0007】(ウ) 記事の参照関係は記事の中にあるヘッダの情報を利用してゐた。そこに明示的に参照関係が書いてない記事については、関連が深くても独立した木として表示される。ヘッダ情報がなくても、関連した情報が1つの木として表示される手段を提供する必要がある。

【0008】(エ) 参照関係を使った木の中には大きくなりすぎて、複数の話題を含んでいる場合がある。このような場合は、話題毎に木の部分を分解して複数の木として表示する手段を提供する必要がある。

【0009】(オ) 1つの話題を表す木の中でユーザが特に興味のある話題を指定できない。本発明の目的は、これらの課題を満たすニュースリーダを作成することである。

【0010】

【課題を解決するための手段】

(1) 記事の集合を納める記事蓄積装置を持ち、該記事はヘッダ部分と内容部分からなり該記事のヘッダ部分には、その記事を識別するための識別名と、その記事が別の記事を参照して書かれた場合には、参照している記事の識別名を含み、内容部分にはテキストデータを含み、該記事の内容部分を表示する内容表示画面からなる電子掲示板ブラウザにおいて、該記事の間の参照関係を表示する木構造表示画面と該木構造表示画面の上に表示されている該記事をユーザが指定するためのポインティング装置と、該記事蓄積装置内部の記事を調べて参照関係を表す参照関係データを作成する参照関係解析装置、該参照関係解析装置が作成したデータを木構造表示画面に該記事をノードで、参照している関係を参照される記事から参照した記事へのリンクで表示する木構造表示装置と、ユーザによって指定された該ノードの子孫のノードを該木構造表示画面から隠す機能を有する木構造操作装置を持ち、ユーザによって該ポインティング装置で指定された該木構造表示画面のノードに対応する記事の内容

を該内容表示画面に表示することを特徴とする。

【0011】(2) 上記(1)の電子掲示板ブラウザにおいて、複数の子供を持つノードにおいて、子供が終端ノードであるとき、他の子供で終端でないものがあるれば、終端の子供を木から除去する機能からなる話題抽出装置を持つことを特徴とする。

【0012】(3) 上記(1)の電子掲示板ブラウザにおいて、各記事のヘッダ部分に参照を表すデータがないとき、参照関係解析装置が各記事からキーワードの集合を抽出する機能を持ち、キーワードの集合の包含関係を使って記事の間のリンクを張ることを特徴とする。

【0013】(4) 上記(1)の電子掲示板ブラウザにおいて、各記事のヘッダ部分に参照を表すデータがないとき、参照関係解析装置が各記事から他記事を参照する部分に現れやすい予め決められている文字列を検索し、その文字列の近くの文章からキーワードを抽出し、そのキーワードを含む記事を探して参照のリンクを張ることを特徴とする。

【0014】(5) 上記(1)の電子掲示板ブラウザにおいて、各記事に対して記事の内容を表す特徴ベクトルの蓄積し、親子関係のあるノード間で該特徴ベクトルどうしの差があらかじめ決めたある値以上であるとき、該ノード間のリンクを切る話題分割装置を持つことを特徴とする。

【0015】(6) 上記(5)の電子掲示板ブラウザにおいて、該話題分割装置は、各記事の特徴量の差でなく、その記事のリンクに関する近傍の記事の特徴量の加重平均の差分をリンクの切り離しの条件に用いることを特徴とする。

【0016】(7) 上記(5)の電子掲示板ブラウザにおいて、該話題分割装置はリンクを切る前に切る部分を該木構造表示画面で示し、ユーザに切るかどうか聞いて確かめることを特徴とする。

【0017】(8) 上記(1)の電子掲示板ブラウザにおいて、木の一部分を別の記事の子供に付け替える機能を持つ木構造操作装置を持つことを特徴とする。

【0018】(9) 上記(1)の電子掲示板ブラウザにおいて、ユーザがあるノードの集合を指定して、そのノードの子孫となる記事が該記事記憶装置に入力されたら、ユーザにその記事を知らせる記事監視装置を持つことを特徴とする。

【0019】(1) 本発明のブラウザは、ニュースグループ内の記事の集まりを、参照関係解析装置が調べた記事間の参照関係をリンクとして木構造で表示するので、記事の集まりは、木構造表示装置により木構造表示画面の上で、1つの話のまとまり毎に別の木に分割されて表示され、ユーザは、ポインティング装置により、木のノードを指定し、その内容を内容表示画面に表示できるので、ユーザは、木のルート(根)に近い部分のノードの内容だけを調べれば、その木で何が議論されているのか

わかり、興味があれば、その木の中のノードの内容を一つずつポインティング装置で指定し、内容表示画面に表示し、また、ルート近くの記事の内容を見て興味がなければ、木ごと飛ばして読むことができ、また、木が多くの記事からなっていて、1つの木の内部にユーザの興味のない話題がある可能性がある場合は、ユーザは幾つかのノードを調べ、興味がなければ、木構造操作装置により、そのノードの子孫を画面から消すことにより、ユーザが把握しやすい大きさの木で1つの話のまとまりが表示され、課題(ア)を解決する。

【0020】(2)本発明のブラウザは、一つのノードを複数の記事が参照していて、話題の枝分かれが多いとき、話題抽出装置は、ノードから枝別れしてすぐ議論が終わっているノードを取り去ることが出来、1つの木のノードの個数を減らすことが出来るので、ユーザは1つの木の内容を知るときに少ない個数の記事だけを参照すればよく、より課題(ア)と(イ)の解決に貢献する。

【0021】(3)本発明のブラウザは、参照関係解析装置は、各記事からキーワードの集合を集め、2つの記事の間に関係があるかどうかを調べるのに、キーワードの包含関係を調べるので、記事の中に明示的に参照関係が書いてなくとも記事の集合を木にまとめることができ、より課題(ア)と(ウ)の解決に貢献する。

【0022】(4)本発明のブラウザは、参照関係解析装置は、各記事から予め決められている他の記事を参照するときに現れやすい、「述べられている」や「で言っている」などの語の近くの文章からキーワードを抽出し、それらのキーワードが含まれる記事を探し、参照しているリンクを張るので、記事の中に明示的に参照関係が書いてなくとも記事の集合を木にまとめることができ、より課題(ア)と(ウ)の解決に貢献する。

【0023】(5)本発明のブラウザは、話題分割装置が、ノードの親子間で特徴ベクトルの差分をとって、それが大きく異なるときは、親子関係のリンクを切って、子供のノードを木として独立させるので、参照関係だけでは1つの話題とみなされる特徴の異なる2つの話題を分割することができ、木のルートの付近の記事だけ調べることによって木全体で議論されていることを推測することが、より確かになり、課題(ア)と(エ)の解決に貢献する。

【0024】(6)本発明のブラウザは、請求項5のブラウザで、話題分割装置が特徴関数間の差分を計算する処理において、計算するノードの近傍での特徴関数の平均をとってから差分をとるので、ノイズに強い分割が可能であり、より課題(ア)と(エ)の解決に貢献する。

【0025】(7)本発明のブラウザは、請求項5のブラウザの話題分割装置がリンクを切る処理において、リンクを切る基準を満たしたとき、自動的に切るのではなくユーザに切つてよいかどうか確認するので、よりユーザの意図に合う分割が可能になり、より課題(ア)と

(エ)の解決に貢献する。

【0026】(8)本発明のブラウザは、木構造操作装置を備え、ユーザが木の部分木を別のノードの子供に付け替えたり、新しい木として独立させたりすることができ、大きな木を複数の木に分割して整理することができ、より課題(ア)と(エ)の解決に貢献する。

【0027】(9)本発明のブラウザは、ユーザが興味のある話題をノードで指定し、そのノードで指定された話題に属する記事が記事蓄積装置に入力されたとき、それをユーザに知らせる記事監視装置を持ち、ユーザは関心のある話題をすぐに入手でき、より課題(ア)と(オ)の解決に貢献する。

【0028】

【発明の実施の形態】実施例を用いて本発明を詳細に説明する。図1は、本発明の構成図である。本発明の機械は、記事1002を蓄積するための記事蓄積装置1001、その中に蓄積されている記事1002の内容を表示する内容表示画面1005にからなる電子掲示板ブラウザに次の装置が追加されている。

【0029】すなわち、(1)記事の間の参照関係を表示するための木構造表示画面1003、(2)記事蓄積装置1001内の複数の記事の間の参照関係を解析する参照関係解析装置1008、(3)木構造表示装置に表示された木のノードを指定するためのポインティングデバイス1006、と、次の(5)から(9)までの装置からなる記事データ操作装置1004(5)参照関係解析装置1008が解析した結果の参照関係データ1007を木構造表示画面1003に表示する木構造表示装置1009、(6)木のノードを表示画面から見えなくする機能と木の一部分を別の部分に移動する機能を持つ木構造操作装置1010、(7)多くの枝わかれしているノードの子供のノードで話題がそこで終端しているノードを取り去る話題抽出装置1011、(8)大きな枝の中の独立している話題を分割して別の木にする話題分割装置1012、(9)ユーザが指定したノードから派生する話題に属する記事が記事蓄積装置に入力されたとき、それをユーザに知らせる記事監視装置1013、である。

【0030】図2は、記事蓄積装置1001に蓄積されている記事の集合を解析して作られる記事間の参照関係を表現するデータである。このデータは、3つのフィールド「元の記事」と「左の記事を参照している記事の集合」と「表示フラグ」からなるテーブルである。図2の第1行目は、記事2と記事3は記事1を参照していることを表している。また第3行目は、記事3はどの記事からも参照されていないことを表している。

【0031】図3は、記事1002のデータ形式を表す図である。記事はヘッダ部分と内容部分に別れ、ヘッダ部分には、その記事の識別名と参照する記事の識別名が記録されている。図2では、記事の識別名は

<19950107102301M94A>;

であり、参照する記事の識別名は、
 <19941230152308N2C>
 という文字列である。

【0032】図4は、記事蓄積装置1001に含まれる記事の集まりを解析して、参照関係データを作成する記事参照関係解析装置の動作を示す。まず、処理4001で、参照関係のデータを空にする。次に処理4002で、記事蓄積装置1001に入っているすべての記事Xに対して、処理4003の登録作業を行い、参照関係データの第1フィールドに各記事Xの識別名をもつ行を作成する。このとき「表示フラグ」はすべてTRUEにしておく。次に、処理4004で、記事蓄積装置1001に入っているすべての記事Xに対して処理4005以下の処理を行い、記事の間の参照関係のデータを作成する。まず処理4005で、Xのヘッダに他を参照している情報が入っているかどうかを調べる。もし入っているなら、処理4006で、それをYとし、Yを第1フィールドに持つ行を参照関係データから見つけたし、処理4007で、Yの第2フィールドにXを加える。

【0033】図5は、参照関係データから図1の木構造表示画面1003に木構造を表示する装置である木構造表示装置1009の動作を示すアルゴリズムを示している。まず、処理5001で、変数TOPSにどの記事も参照していない記事の集合を入れる。この記事の集合が表示される木の集合の各々の要素のルートである。処理5002において、最初の木をルートを表示するXY座標をそれぞれ変数LEFTとTOPに設定する。この変数は、このアルゴリズムの中で使うサブルーティン $DISPLAY(NODE, X, Y)$ の中から参照や設定ができる大域変数であるとする。 $DISPLAY(NODE, X, Y)$ については次の図6で説明する。

【0034】次に、処理5003で、変数TOPSに属する記事の1つ1つについて処理5004以下の処理を繰り返し、その記事を参照していて、かつ、表示フラグがTRUEである記事の木を木構造表示装置1003に描く。変数TOPS内の各記事NODEについては、まず、処理5004で、NODEの表示フラグがTRUEかどうか調べ、もしTRUEならば表示するのであるから、処理5005でサブルーティン呼び出し $DISPLAY(NODE, LEFT, TOP)$

をして、NODEから参照のリンクを使って到達可能な記事を座標 (LEFT, TOP) を左上として描く。画面の座標は、左上を (0, 0) とし、X は左方向に大きくなり、Y は下方向に大きくなるように座標系をとってあるとする。 $DISPLAY(NODE, LEFT, TOP)$ の中で、TOP は、NODE からなる木を表示したとき最も下に表示されたノードのY座標を保持するように更新される。次に、処理5006で、次の木を描くとき、上の木とノードを描く領域が一致しないように、

$BOTTOM = BOTTOM + DELTA_Y$

で間隔をとっている。

【0035】図6は、図5のアルゴリズムの中で使用されるサブルーティン $DISPLAY$ のアルゴリズムを示してい

る。このサブルーティンが

$DISPLAY(NODE, X, Y)$

の形で呼び出されたときの処理を記述している。NODE は、記事を表し、X, YはX, Y座標を表している。まず、処理6001で指定されたノード NODE を、(X, Y)を左上として矩形で描く。このとき、矩形の右下座標を(X1, Y1)とする。次に処理6002で、NODE の子供を描くために X 座標をX1より少し右にとった値を変数X2に設定している。次に処理6003で、子供を描きはじめるY座標 Y2 を Yに設定する。

【0036】次に処理6004で、記事NODE の各子供の記事Cで、表示フラグがTRUEである記事にたいして処理6006から処理6008までを行い、木の描画を完成させる。まず、処理6005で、 $DISPLAY$ を再帰的に呼び出して、Cを(X2, Y2)を左上として描く。次に処理6006で、NODE を表す矩形からCを表す矩形へ線を引く。これが、参照のリンクを表す線である。次に処理6008で、C の後ろにまだNODEの子供で表示フラグがTRUEであるものがあれば、 $Y2 = BOTTOM + DELTA_Y$

で、次の子供を書きはじめるY座標を今描いた木の下座標 BOTTOM に木と木の間隔 DELTA_Y だけ加えた値にする。すべての子供について処理が終わったら、処理6009、処理6010、処理6011で、変数 BOTTOM を $DISPLAY$ が描いた木の一番下の座標に設定する。すなわち、子供の一番したの座標と、NODE 自身の矩形の一番下の座標 Y1 を比較して、より下の方を BOTTOM の値として設定する。

【0037】次に、木構造操作装置1010が指定されたノードを木構造表示画面1003から隠す方法について説明する。図2において、「表示フラグ」は、最初に参照関係データが作られたときは、真を表す値 TRUEが入っている。木構造操作装置1010が、ノードを木構造表示画面から隠すために用いる。すなわち、木構造操作装置1010は、指定されたノードおよびその子孫のノードを木構造表示画面から隠すときは、この「表示フラグ」のフィールドに偽を表す値FALSEを入れる。このとき、次に図5および図6で説明した木構造表示装置1009は、このフィールドの値を調べてFALSEなら、そのノードおよび子孫のノードを表示しないようになっており、そのフィールドにFALSE を入れて木構造表示装置1009で木を再表示すると指定されたノードおよびその子孫のノードは木構造表示画面から隠される。

【0038】図7は、話の枝葉の部分の木の形を示している。ノード7003は話の枝葉と解釈する。ノード7001, 7002 からの話は、ノード7004, 7007 以下のノードで展開されていると見なして、話の本筋を抽出する場合、ノード7003は除去する。

【0039】図8は、話の本筋だけを抽出するアルゴリズムを記述している。処理8001ですべてのノード Xについて処理8002以下の木の中から話題を抽出する処理を繰

り返す。まず、処理8002で、X のすべての子供を調べ、その子供の中にさらに子供を持つものがあるかどうか調べる。もし、あるなら変数 HAS_NEXT を TRUE にし、そうでなければ FALSE にする。次に処理 8003 でXの子供のノード Y について処理8004以下の繰り返す。まず、処理8004で、Yが子供をもたないなら、処理8005で、他に子供を持つ子供があるかどうかを調べ、もしあれば、Yは、そのあと議論が続いていないノードであるから処理8006でYを除去する。もし、処理8005でHAS_NEXTがFALSEならば、他の子供も議論が続いておらず、どの記事を消してよいか基準がないので、Yは消さずに残しておく。

【0040】図9は、記事の中でキーワードを集めて、キーワードの集合間の包含関係を使って記事間のリンクを設定するアルゴリズムを記述している。まず処理9001で、すべての記事からキーワードを抽出し、各記事に対応つけておく。次に処理9002と処理9003ですべての記事の組み合わせ(X, Y)に対して、処理9004で、記事Xのキーワードの集合Keywords(X)が記事Yのキーワードの集合Keywords(Y)に殆ど含まれているなら、処理9005でYからXにリンクを張る。ここで用語「殆ど含まれる」は、次のように定義する。ただし、各キーワードには、重要度を表す重みが付けられているとする。

【0041】A, Bをキーワードの集合とし、 $\text{Diff}(B, A)$ = 集合の差 (B - A) に含まれるキーワードの重みの合計とすると、用語「殆ど含まれる」を A は殆ど B に含まれる $\Leftrightarrow \text{Diff}(B, A) < \epsilon$; ある定数で定義する。

【0042】本アルゴリズムでリンクを付けた場合は、参照関係のサイクル(記事Aから参照のリンクを辿って再び記事Aに戻るリンクの列)ができる可能性があるが、そのサイクルは木の表示装置がサイクルを見つけたときに、切ることにする。

【0043】図10は、記事の中で予め登録された別の記事を参照するときに現れやすい語を検索し、その語付近に現れたキーワードの集合を抽出し、そのキーワードの集合を含む記事へリンクを設定するアルゴリズムを記述している。

【0044】ここでは、予め登録された別の記事を参照するときに現れやすい語は、「述べ」、「言う」、「言い」などを登録してあるとする。

【0045】また、文中に現れる語の近くとは、ある正の整数Nを決めておいて、その語の始まりの位置の前後N文字以内としておく。まず、処理10001で、各記事Xについて処理10002以下の処理を繰り返しXが参照している記事を見つけてリンクを張る。まず処理10002で、予め登録されている他の記事を参照するとき現れやすい語がXの中に現れるかどうか調べ、現れたらその語の出現の前後N文字の中にキーワードが現れるかどうか調べ、現れたキーワードを集めておく。次に処理10003で、キーワードが十分多く集まったかどうか判定する。これは予

め決められた数以上キーワードが集まったら、十分多く集まったと判定する。この判定で十分多くのキーワードが集まったと判定されたら、処理10004で、記事蓄積装置10001の中から、処理10002で見つかったキーワードをすべて含む記事を探し出す。処理10005で、それらのキーワードを含む記事が見つかったかどうか判定し、見つかったなら記事Xは、その記事を参照していることを示すリンクを張る。処理10003で、元の記事から十分多くのキーワードが集まらなかったと判定したら、参照関係のリンクを張るためには十分な精度のキーワードがないと考えて、Xは何も参照してなかったことにする。

【0046】図11は、親子のノード間で記事の特徴ベクトルの差分をとり、木の分割を行うアルゴリズムを示している。まず、処理11001で、各ノード X に対して特徴ベクトル

$$C(X) = \langle C_1(X), \dots, C_n(X) \rangle$$

を計算する。ここで $C_i(X)$ $i = 1, \dots, n$ は、X のある特徴を表すスカラーである。特徴関数は用途に応じて、設定することが必要であるが、例えば、次のような特徴関数を使えばよい。

【0047】 $C(X) = \langle \text{分類1のキーワードの得点}(X) / X\text{の文字数}, \text{分類2のキーワードの得点}(X) / X\text{の文字数}, \text{逆説接続詞の割合}(X) \rangle$

ここで、分類1のキーワードの得点(X)

記事 X に含まれる分類1に分類されるキーワードの重みの合計

分類2のキーワードの得点(X)

記事 X に含まれる分類2に分類されるキーワードの重みの合計

逆説接続詞の割合(X)

記事 X に含まれる逆説接続詞の個数の全接続詞に対する割合とする。上の特徴関数では、記事の分野が変われば、ベクトルの方向が大きく変わることになる。上記のように特徴関数を計算したら、次に処理11002で、各ノードXに対して、処理11003以下の処理を行って、話題の切れ目を探し、リンクを切断する処理を行う。まず、処理11003でXの親を Y とする。次に処理 11004 でXとYの特徴量の値の差をとり、ある定数 Alpha と比較する。

【0048】 $\text{DIFF}(C(X), C(Y)) > \text{Alpha}$

ここで、DIFFは、二つのベクトルの差をスカラーで表すための関数である。次のようにとっておけばよい。

【0049】

$$\text{DIFF}(A, B) = |A_1 - B_1| + \dots + |A_n - B_n|$$

ここで、 $A = (A_1, \dots, A_n)$, $B = (B_1, \dots, B_n)$ 、 $|u|$ は、uの絶対値とする差がその定数より大きければ、処理11005で、XをYから切り離す。

【0050】図12は、差分による木の分割を行うとき、ノイズデータにより話題が変化していないにもかかわらず切れてしまうのを防ぐために木の親子関係のリンクでの近傍で特徴量の移動平均をとるアルゴリズムを示

している。ここでは、 n を0以上の整数として、距離 n 以内の移動平均を求めている。処理12001で、各記事 X について、処理12002以下の処理を行い、 X に対して移動平均を計算していく。各 X に対して、処理12002で子供の方向にリンクを辿り、 n 回以内で到達できるノードを集め、それを変数 NEIGHBOR に入れる。自分自身は0回で到達できるので、NEIGHBORの中には X は入っている。処理12003で、NEIGHBORの中の要素数を m とする。 X がNEIGHBORの要素なので、 m は、1以上である。処理12004で、NEIGHBOR内の記事の特徴関数の合計を計算し、それを S とし、記事 X の特性の移動平均として S/m をノード X に記録する。

【0051】図13は、実際の木における距離2の移動平均の計算例を示している。記事12003からリンクを2回以内辿って到達できる記事は、13003, 13004, 13005, 13007, 13008, 13009である。従って、これらの特徴関数の合計を記事の個数6で割ったものが記事13003の移動平均であるから、記事13003の移動平均 $= (70 + 50 + 20 + 40 + 27 + 33) / 6 = 40$ となる。

【0052】図14は、親子のノード間で差分が大きかった箇所のノードを色を変えて表示するアルゴリズムを示している。これにより話題が切れる可能性がある部分を表示して、ユーザに切ってよいかどうか聞けば、より分割がユーザに分かりやすいものになる。まず、処理14001で、各記事 X について、処理14002以下の処理を繰り返し、各 X に色を付けていく。まず、処理14002では X の特徴量とその親の特徴量の差分をとる、このとき X に親がなければ、差分は0としておく。次に処理14003でこの差分がある定数Alphaより大きいかどうか調べて、大きければ処理14004で、 X を赤い色にする。大きくなければ、処理14005で X を白い色にする。このアルゴリズムを適用した結果赤くなった記事は親の記事から切り離され、新たに木のルートになる可能性のある記事である。

【0053】図15は、木の一部分を別の記事の子供に付け替えるアルゴリズムを表している。ここでは、ノード A をノード B の子供にする方法を述べている。まず、処理15001で、参照関係データ1007の第2フィールドを調べ、記事 A を含むものがあるかどうかを調べる。見つければ、その行を R_A とし、見つからなければ R_A は-1とする。次に処理15002で R_A が-1かどうかを調べ、もし、-1でなければ、処理15003で R_A 行の第2フィールドから A を取り去る。次に処理15004で第1フィールドが B である行があるかどうか調べる。もしあれば、その行を R_B とし、なければ新たに行を作成して、それを R_B とする。最後に処理15005で、記事 A を R_B 行の第2フィールドに付け加えれば、 A は、 B の子供として登録される。

【0054】図16は、ブラウザのノードを指定し、記事蓄積装置に新しい記事 X が入ってきたとき、指定された記事の子孫になっていたら、ユーザにその記事を知ら

せるアルゴリズムを示している。このアルゴリズムでは入ってくる記事の順番は、必ずしも参照の順番にくるとは仮定していない。すなわち、ある記事 A を参照した記事 B が入ってきた後、 A が入ってくることも有り得ると仮定している。この過程は現実のニュースシステムで起こり得る。まず、処理16001で、ユーザが監視を指示した記事の子孫に対して監視するマークを付ける。ここで、ある記事の子孫とは、その記事およびその記事からリンクを子供の方向に1回以上辿って到達できる記事のことである。次に処理16002で、記事の集合を格納するための変数NEWを空にする。このアルゴリズムが終了したあと、NEWに溜まった記事がユーザが指定したノードの子孫で新しく入力されたものである。次に処理16003で、新しく記事蓄積装置に入力された記事を X とする。処理16004で、現在記事蓄積装置にある X 以外の記事 Y について X との親子関係を調べていく。まず、処理16005で、 X が Y の子供であるときは、処理16006で Y から X にリンクを張り、処理16007で Y に監視マークがついているかどうか調べ、もし付いていれば、処理16008で、 X および X の子孫で監視マークのついていないものに監視マークを付けて、それらの記事を変数NEWに入れる。処理16005で X が Y の子供かどうかしらべた後、今度は処理16009で、 Y が X の子供かどうか調べる。もし、そうなら処理16010で X から Y へリンクを張り、処理16011で X に監視マークが付いていて Y に付いていない場合は、処理16012で Y および Y の子孫で監視マークのついていないものに監視マークを付けて、それらの記事を変数NEWに入れる。すべての記事について上記の処理を終えたあと、変数NEWには、監視が指定されたノードの子孫でまだマークがついていなかったものが入っているため、処理16013で、それをユーザに知らせて処理を追える。

【0055】

【発明の効果】本発明によれば、多くの記事の集まりをユーザが一つずつ内容を調べなくても、話のまとまり毎に分類してみることができる。

【図面の簡単な説明】

【図1】本発明実施例の構成図。

【図2】参照関係データの形式を表す図。

【図3】記事のデータ形式。

【図4】参照関係解析装置のアルゴリズムを表すPAD図。

【図5】複数の木の表示を制御するアルゴリズムを表すPAD図。

【図6】1つの木を表示するアルゴリズムを表すPAD図。

【図7】木から話題を取り出すとき捨てられる記事を表す図。

【図8】木から話題を抽出する手続きのPAD図。

【図9】記事のキーワードを収集して記事間の参照関係を決める処理のPAD図。

【図10】他の記事を参照している部分を発見する手続きのPAD図。

【図11】話題の切れ目でリンクを切る手続きのPAD図。

【図12】木で特徴量の移動平均をとる手続きのPAD図。

【図13】木における移動平均の計算例を表す図。

【図14】木において話題の切れ目の記事の色を変える手続きのPAD図。

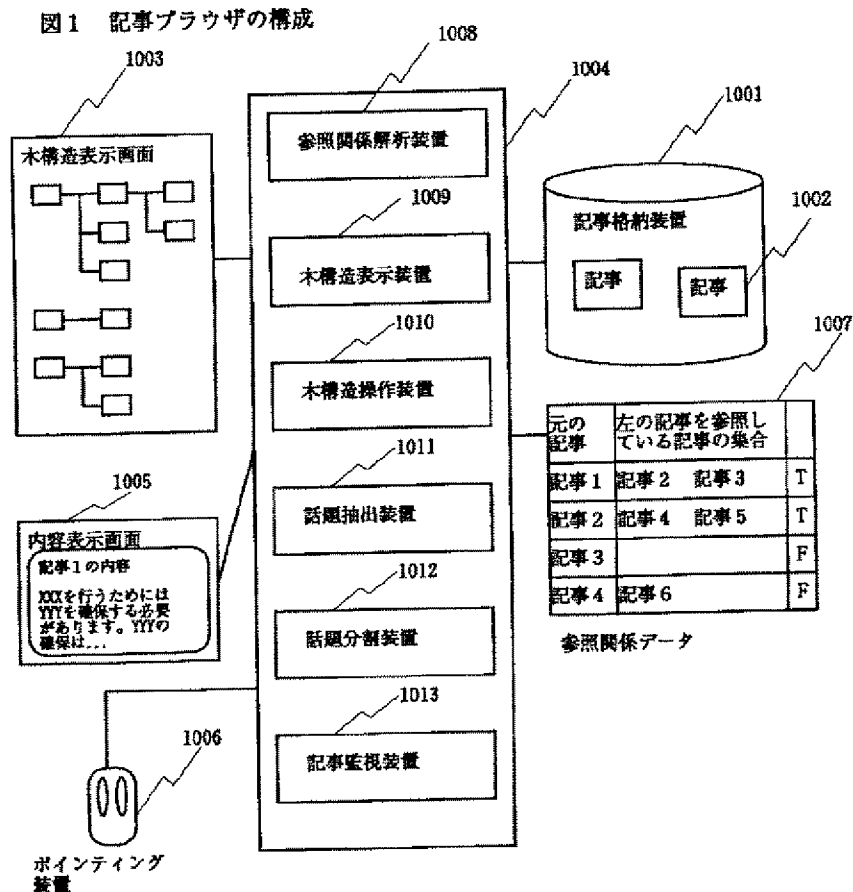
【図15】木の一部分をあるノードの子供にする手続きのPAD図。

【図16】指定された記事の子孫が入力されたかどうか監視する手続きのPAD図。

【符号の説明】

1001 ... 記事記憶装置
1002 ... 記事
1003 ... 木構造表示画面
1004 ... 記事データ操作装置
1005 ... 記事内容表示画面
1006 ... ポインティング装置
1007 ... 参照関係データ
1008 ... 参照関係解析装置
1009 ... 木構造表示装置
1010 ... 木構造操作装置
1011 ... 話題抽出装置
1012 ... 話題分割装置
1013 ... 記事監視装置

【図1】



【図2】

図2 参照関係データ

元の記事	左の記事を参照している記事の集合	表示フラグ
記事1	記事2 記事3	TRUE
記事2	記事4 記事5	TRUE
記事3		FALSE
記事4	記事6	FALSE

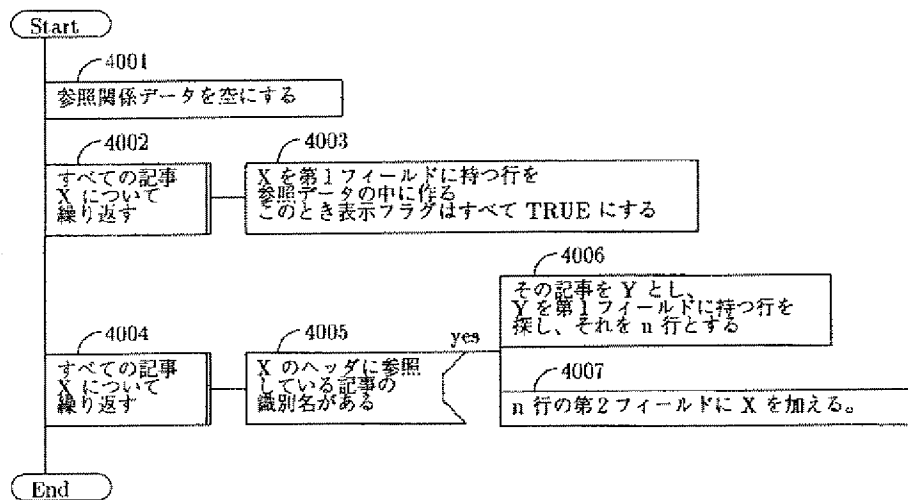
【図3】

図3 記事のデータ形式

ヘッダ:
 記事のID: <19960107102301M94A>
 参照する記事のID: <19941230152308N2C>
 内容:
 前の記事で述べられているXXXの形式について
 コメントがあります。その記事ではXXXは、
 YYYとZZZからなると書いてありますが
 実際は、UUUである場合もあり、.....
 :
 :

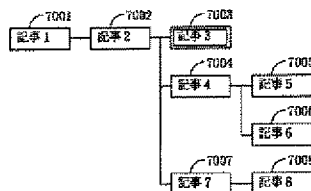
【図4】

図4 参照関係解析アルゴリズム



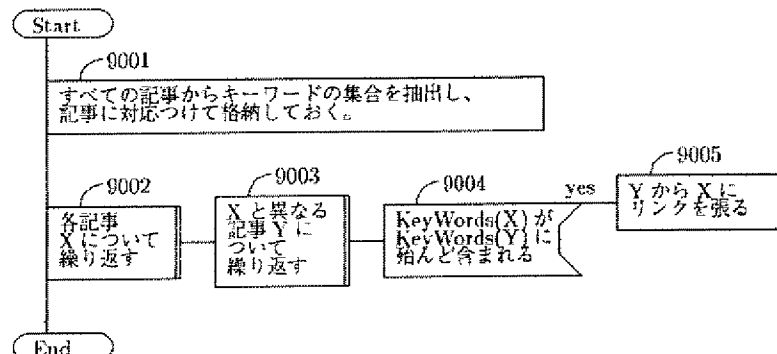
【図7】

図7 記事の木における話題



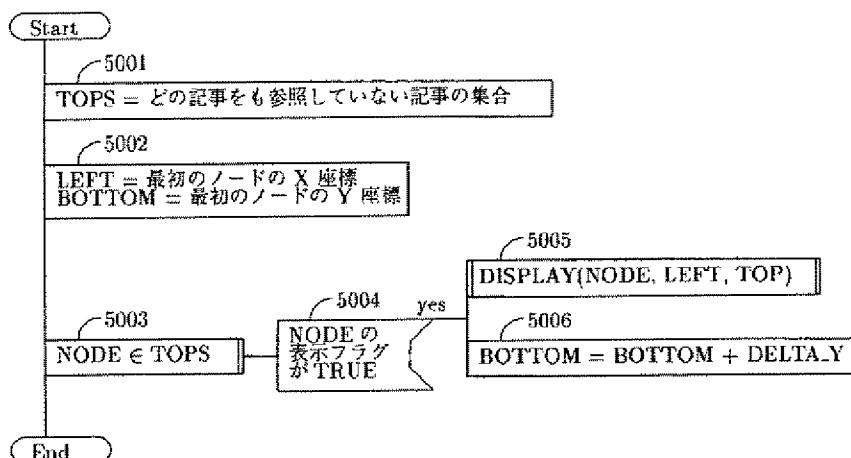
【図9】

図9 キーワード包含関係でリンクを張るアルゴリズム



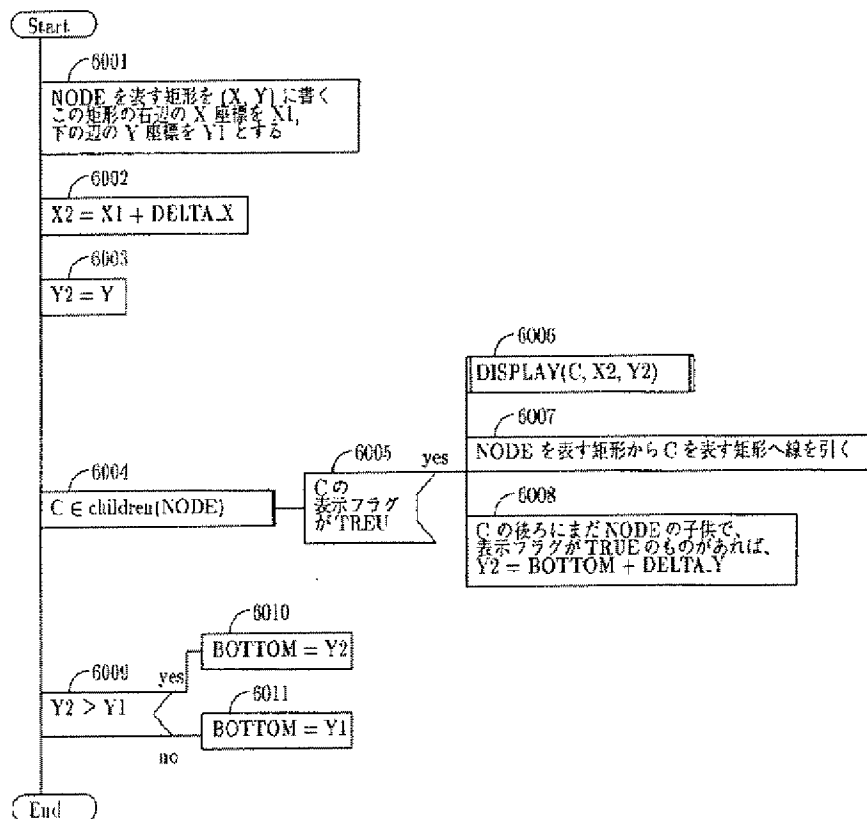
【図5】

図5 木の表示アルゴリズム



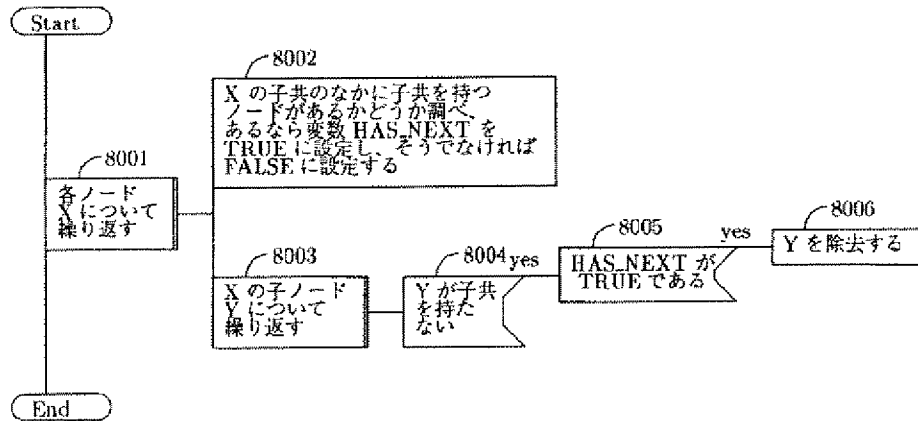
【図6】

図6 DISPLAY(NODE, X, Y)



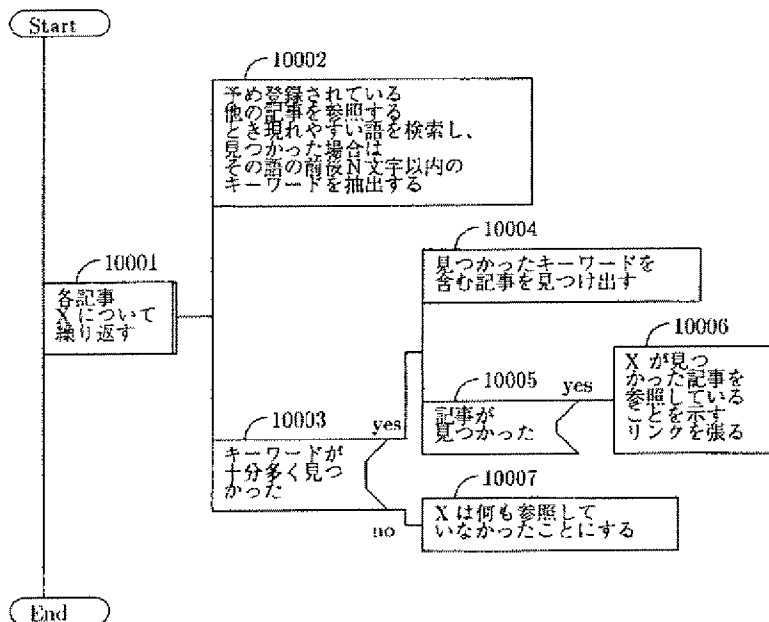
【図8】

図8 話題の抽出アルゴリズム



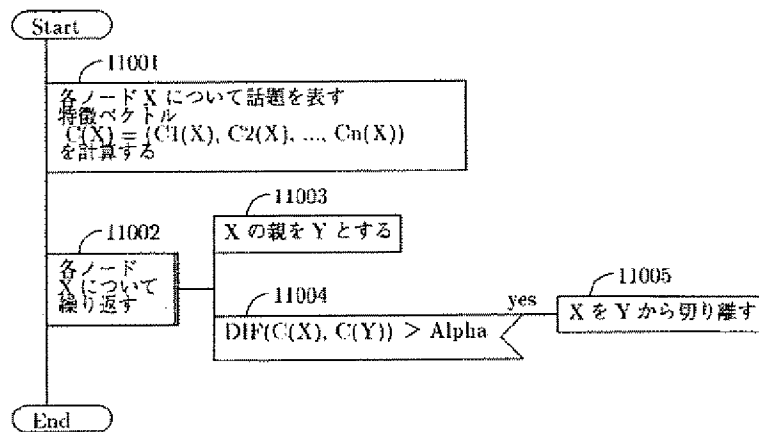
【図10】

図10 他記事参照部分の推定アルゴリズム



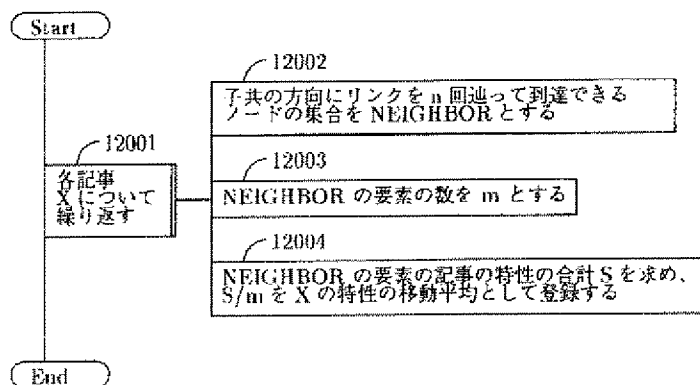
【図11】

図 11 話題の切れ目でリンクを切るアルゴリズム



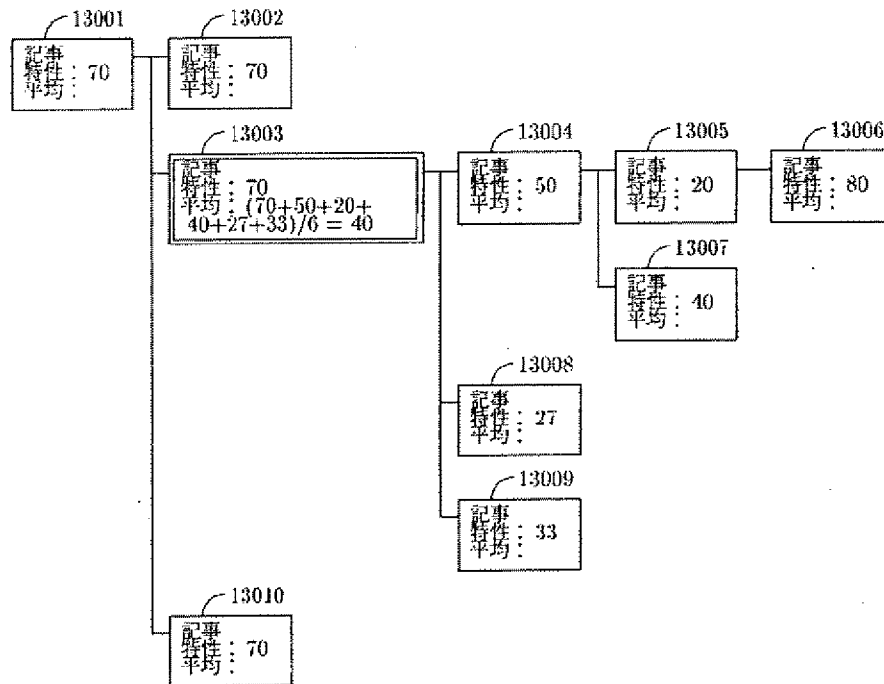
【図12】

図 12 木で移動平均をとるアルゴリズム



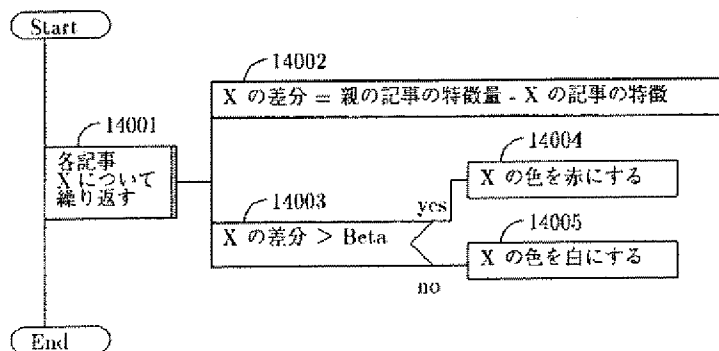
【図13】

図 13 木における移動平均の例



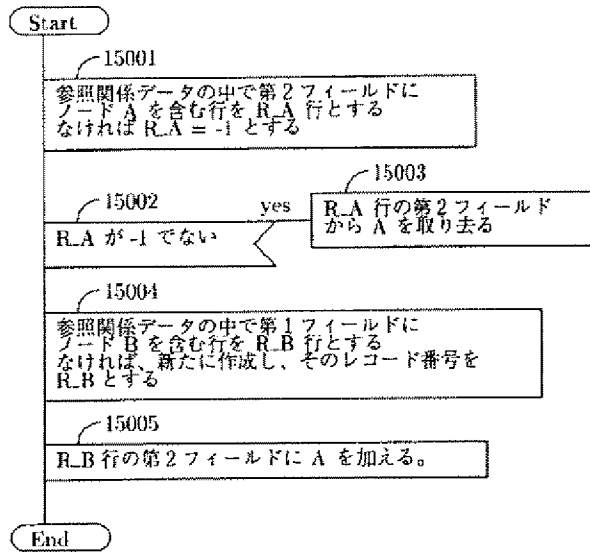
【図14】

図 14 木の枝が切れる可能性を表示するアルゴリズム



【図15】

図 15 木の一部分の移動



【図16】

図 16 子孫の入力監視アルゴリズム

